

# Reconocimiento de Patrones 2018

## IIE, Facultad de Ingeniería, UdelaR

### Práctico 2

Entrega: martes 9 de octubre (eva); jueves 11 de octubre en clase (versión papel)

## Decisión bayesiana

### Ejercicio 1

En un problema de clasificación sea  $\omega_{\max}(\mathbf{x})$  la clase para la cual

$$P(\omega_{\max}|\mathbf{x}) \geq P(\omega_i|\mathbf{x}) \quad \forall i = 1, \dots, c.$$

1. Mostrar que  $P(\omega_{\max}|\mathbf{x}) \geq \frac{1}{c}$ .
2. Mostrar que para la regla de decisión de mínimo error de Bayes, la probabilidad media de error vale

$$P(\text{error}) = 1 - \int P(\omega_{\max}|\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

3. Usar estos resultados para mostrar que  $P(\text{error}) \leq (c-1)/c$ .
4. Describir en qué situación se cumple  $P(\text{error}) = (c-1)/c$ .

### Ejercicio 2

Se considera el test de hipótesis simple binario, sobre una medida escalar  $X$ :

$$\mathcal{H}_0 : X \sim N(0, \sigma^2)$$

$$\mathcal{H}_1 : X \sim N(\mu, \sigma^2)$$

Los valores  $\mu > 0$  y  $\sigma^2$  son conocidos.

1. Calcular las probabilidades de falsa alarma y de detección,  $P_{FA}$  y  $P_D$ , suponiendo que la regla de decisión es:

Decidir  $\mathcal{H}_1$  si  $X > \gamma$ ;  $\mathcal{H}_0$  si no.

El resultado se expresará por medio de la función  $Q(\lambda) = \int_{\lambda}^{+\infty} \exp(-t^2/2)/\sqrt{2\pi} dt$ . Graficar las densidades indicando las áreas de error.

2. Expresar  $P_D$  como una función de  $P_{FA}$ , y graficar la curva correspondiente. ¿Qué representa esta curva?
3. Diseñar una regla de decisión de Neyman-Pearson con restricción  $P_{FA} \leq \alpha$ .

# Discriminantes lineales

## Ejercicio 3: Perceptron

Inspirado por el modelo de neurona propuesto por McCulloch-Pitts (MCP), Frank Rosenblatt propuso un algoritmo capaz de aprender automáticamente los coeficientes de un vector  $\omega$ , que al ser multiplicado por el vector de entrada, permitía decidir si la muestra pertenecía a una clase o no.

1. Explique brevemente la estrategia seguida para determinar el vector  $\omega$
2. Implementelo y muestre que puede separar correctamente las clases que se encuentran en el archivo `dataset1Pr2.csv`. Para ello dibuje los patrones de cada una de las clases con un color distinto y muestre el límite de decisión encontrado.
3. Utilice el mismo algoritmo para separar el siguiente conjunto de datos:

$x_0$	$x_1$	clase
0	0	1
0	1	0
1	0	0
1	1	1

Comente los resultados obtenidos.

## Ejercicio 4: Criterio de Fisher

Fisher propuso como criterio de separación entre dos clases, maximizar el siguiente cociente

$$J(\omega) = \frac{(\omega^T \mu_1 - \omega^T \mu_2)^2}{\omega^T \Sigma_1 \omega + \omega^T \Sigma_2 \omega}$$

siendo  $\mu_1$ ,  $\Sigma_1$ ,  $\mu_2$  y  $\Sigma_2$  las medias y las matrices de covarianza de las clases 1 y 2 respectivamente.

1. Para el conjunto de datos `dataset1Pr2.csv`, encontrar los coeficientes de  $\omega$  óptimos
2. Mostrar la dirección obtenida junto con los datos
3. Implemente el clasificador y reporte los resultados

## Ejercicio 5: Regresión logística

En este modelo se asume que la probabilidad a posteriori de pertenencia a la clase positiva puede ser escrita de la siguiente forma:

$$P(Y = 1 | \mathbf{x}_i; \omega) = \sigma(\mathbf{x}_i; \omega) = \frac{1}{1 + \exp\left(-\left(\omega_0 + \sum_{j=1}^k \omega_{ij} x_{ij}\right)\right)}$$

donde  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})$  es el  $i$ -ésimo vector de características y  $k$  es el número de características.

1. Escribir la log verosimilitud  $l(\omega | \mathbf{X}, \mathbf{y})$  para este modelo en función de la función  $\sigma$  definida anteriormente.  $\mathbf{X}$  es una matriz de  $n \times k$  e  $\mathbf{y}$  es un vector que contiene las  $n$  etiquetas.
2. Mostrar que el gradiente de la log verosimilitud con respecto a  $\omega_j$  está dado por

$$\frac{\partial}{\partial \omega_j} l(\omega | \mathbf{X}, \mathbf{y}) = \sum_{i=1}^n x_{ij} (y_i - \sigma(\mathbf{x}_i; \omega))$$

donde por conveniencia se agrega una dimensión, asumiendo que  $x_{i0} = 1$ . De esta forma, la expresión es válida para  $\omega_0$  también.

3. Implementar un clasificador basado en regresión logística que utilice como método de optimización descenso por gradiente. Utilizarlo para separar el conjunto de datos `dataset2Pr2.csv`.

## Estimación no paramétrica

### Ejercicio 6: Ventanas de Parzen

En este ejercicio se utilizará la base de datos Iris para clasificación de flores disponible en la página del curso. Considere sólo el largo y el ancho del pétalo como características.

1. Grafique la distribución de los patrones de entrenamiento en función de las características.
2. Escriba un programa que estime las densidades de cada clase en una grilla de 50x50 puntos. Asegúrese que la grilla cubra el rango de valores de los patrones de entrenamiento de la clase. Utilice un kernel gaussiano de la forma

$$K(\mathbf{x}, \mathbf{Z}_i^m) = \frac{1}{2\pi^{d/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{Z}_i^m)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{Z}_i^m)}$$

La matriz  $\boldsymbol{\Sigma}$  se considerará de la forma  $\boldsymbol{\Sigma} = h^2\boldsymbol{\Sigma}_c$ , siendo  $\boldsymbol{\Sigma}_c$  la matriz de covarianza de los patrones de entrenamiento de la clase y  $h$  el parámetro que controla el ancho de la ventana.

3. Varíe el ancho de la ventana, grafique las estimaciones obtenidas y establezca un rango de valores razonable para realizar la estimación.